

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 61-016348

(43)Date of publication of application : 24.01.1986

(51)Int.Cl.

G06F 12/08

(21)Application number : 59-136621

(71)Applicant : MITSUBISHI ELECTRIC CORP

(22)Date of filing : 03.07.1984

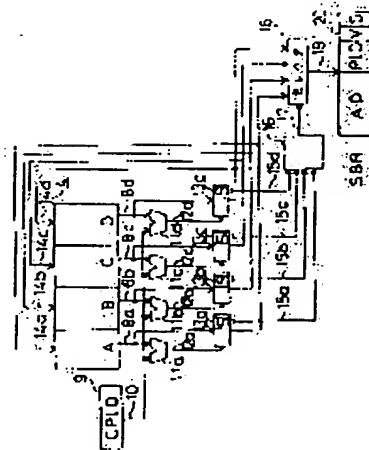
(72)Inventor : NOJI TAMOTSU

(54) BUFFER MEMORY DEVICE

(57)Abstract:

PURPOSE: To reduce the overhead for expulsion of old blocks from a buffer memory at a block replacement time by forcecasting a block in the buffer memory which is considered to be unnecessary in the proppression of a program and preparing for expulsion of these blocks.

CONSTITUTION: As an optional desired program is executed, an address array AA4 is accessed, and contents of its lines AWD are read out to AA output signal lines 8aW8d, and program identifiers (PID) in their entries and a program identifier (CPID) of the present executing program on a CPID output signal line 10 are compared with each other by comparators 11aW11d. S flags in AA data registers 13aW13d are set or reset by outputs of comparators 11aW11d. An S flag detecting circuit 16 generates the signal which selects reset registers out of AA data registers 13aW13d, and this signal is sent to a selector 18.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

⑨ 日本国特許庁(JP)

⑩ 特許出願公開

⑫ 公開特許公報(A) 昭61-16348

⑬ Int. Cl.⁴

G 06 F 12/08

識別記号

庁内整理番号

8219-5B

⑭ 公開 昭和61年(1986)1月24日

審査請求 未請求 発明の数 1 (全4頁)

⑮ 発明の名称 バッファ・メモリ装置

⑯ 特 願 昭59-136621

⑰ 出 願 昭59(1984)7月3日

⑱ 発 明 者 野 地 保 鎌倉市上町屋325番地 三菱電機株式会社計算機製作所内

⑲ 出 願 人 三菱電機株式会社 東京都千代田区丸の内2丁目2番3号

⑳ 代 理 人 弁理士 曾我 道照 外3名

明 細 書

1 発明の名称

バッファ・メモリ装置

2 特許請求の範囲

(1) 主メモリ装置に記憶されている遠数個のブロックに分けられたデータの所要部分の写しをブロック単位で格納し、演算処理装置によつてアクセスされるバッファ・メモリ装置において、任意のブロック内のデータが上記バッファ・メモリ装置に格納されているか否かを指示するエントリ情報を保持するアドレスアレイが設けられ、上記エントリ情報には、少なくとも、上記バッファ・メモリ装置内の対応するブロックにおけるデータが属しているプログラムを識別するための情報、および、当該データを含むブロックが追い出し可能であるか否かを指示するための情報が含まれていることを特徴とするバッファ・メモリ装置。

(2) 上記エントリ情報には、上記バッファ・メモリ装置内の対応するブロックにおけるデータの有効性を指示するための情報が更に含まれている

ことを特徴とする特許請求の範囲第1項記載のバッファ・メモリ装置。

3 発明の詳細な説明

〔発明の技術分野〕

この発明はバッファ・メモリ装置に関するものであり、特に、追い出すべきブロックを予測しておき、ブロック交換のさいのオーバーヘッドを軽減できるようにされたバッファ・メモリ装置に関するものである。

〔従来技術〕

電子計算機を中心とする情報処理システムにおいては、大容量の主メモリ装置(MM)と演算処理装置(BXU)との間に、記憶容量はMMより小さいが、MMよりも高速にアクセスすることのできるバッファ・メモリ装置(HSB、キャッシュと呼ばれることもある)を設け、MMに記憶されている情報の一部の写しを予めHSBに格納しておき、BXUは、通常は、このHSBだけをアクセスすることにより情報処理の高速化がはかられている。BXUによつて実行されているプログラムからみた

ときには、このバッファ・メモリ装置またはキャッシュはトランスペアレントなものであり、プログラムは、これを直接的にアクセスすることができないようになってるのが普通である。このようなキャッシュの一典型として、ストア・イン方式のものが知られており、その構成例が第1図に示されている。

この第1図において、(1)はEXU、(2)はHSB、(3)はMM、(4)はアドレス・アレイ(AA)であつて、HSB(2)にMM(3)のデータの写しが格納されているか否かについてのアドレス情報を保持するためのもの、(5)はHSBバス線であつて、HSB(2)をアクセスするアドレス情報とEXU(1)からのストアデータやHSB(2)からのフェッチデータを伝送するためのもの、(6)はHSB(2)とMM(3)との間で情報伝送を行うためのMMバス線、そして、(7)はHSB(2)をアクセスするHSBアドレス線である。

このような装置においては、通常は、MM(3)に記憶されている情報は連続個のブロックに分けられ、各ブロックの写しをHSB(2)に格納するようにされ

るものである。このような場合には、MM(3)内の情報がHSB(2)のいずれの箇所に格納されているかを表わすアドレス情報がAA(4)に格納されている。なお、前記ブロックは数ワード〜数十ワードの大きさのものであり、また、前記AA(4)のアドレス情報は通常はエントリと呼ばれている。

次に、この第1図に示されている装置の動作について説明する。

EXU(1)からのアクセス要求(ストア又はフェッチ)が発生すると、アクセスのためのアドレス、データ情報がEXU(1)からHSBバス線(5)経由でHSB(2)およびAA(4)に転送される。AA(4)では、アクセスすべきデータがHSB(2)に存在するかどうかを照会される。そして、HSB(2)に必要なデータが存在している場合は、HSBアドレス線(7)を介してHSB(2)内の前記必要なデータがアクセス(ストア又はフェッチ)される。

これに対して、必要なデータがHSB(2)に存在しない場合は、HSB(2)とMM(3)との間でMMバス線(6)を介してブロックの交換が行われる。このとき、

HSB(2)に空き領域が存在しない場合はHSB(2)内の最も使用されなかつたブロックがMM(3)へ追い出され、目的とするブロックがMM(3)からHSB(2)へ転送される。このブロック入れ換えは、良く知られているLRU方式により行われる。そして、HSB(2)に空き領域がある場合は、その空き領域に目的とするブロックがMM(3)からHSB(2)へ転送される。その後目的とするブロックのデータがHSBバス線(5)を介してEXU(1)へ転送される。

ところで、通常、プログラムが走行している状態では、HSB(2)に空き領域が存在する確率は低いものである。従つて、通常の動作状態では、HSB(2)内の最も使用されなかつたブロックがMM(3)へ追い出され、目的とするブロックがMM(3)からHSB(2)へ転送されるブロック交換動作が発生する。1ブロックは通常数ワード〜数十ワードから構成されており、また、ブロック交換作業は逐次処理で行われるため、相当なオーバーヘッドが生じ、更にはシステム性能が低下する要因となる。

従来のバッファ・メモリ装置は、以上のように

構成され、動作するものであるために、HSBに必要なデータが存在しない場合には、HSBのブロックをまず追い出してMMから目的とするブロックを転送する逐次型ブロック交換のやり方がとられており、HSBのブロック追い出しのためのオーバーヘッドが生じ、システム性能が低下するという欠点があつた。

〔発明の概要〕

この発明は上記のような従来のものの欠点を除去するためになされたもので、プログラムの進行過程で不要とされるHSBのブロックを予測し、このブロックの追い出し準備をしておくことにより、ブロック交換の必要が生じた時の、古いHSBブロック追い出しのオーバーヘッドを軽減することのできるようにされたバッファ・メモリ装置を提供することを目的としている。

以下、この発明の実施例を図について説明する。第2図はAA(4)のアドレス情報、すなわちエントリの構成を表わすエントリ構成図であり、この中で、ADは実アドレス情報、PIDはプログラム連

別子。Vはそのエントリが有効かどうかを表わす有効性フラグ。Sはそのエントリで指示される領域は、プログラムで使用されておらず、MM(J)への追い出しがいつでも可能な状態にあることを表わすスタンバイフラグである。なお、PIDには、例えばテーブルアドレス変換におけるSTO(セグメント テーブル オリジン)が割当てられるものである。

第J図は、この発明の実施例の要部を示すブロック図であつて、第I図と同一符号は同一又は相当部分を示す。この第J図において、(8a)~(8d)はAA(4)からのAA出力信号線、(9)は現在実行中のプログラム識別子(CPID)を保持するCPIDレジスタ、(10)はCPIDレジスタ(9)から出力されるCPID出力信号線、(11a)~(11d)はAA出力信号線(8a)~(8d)上の信号とCPID出力信号線(10)上の信号とを比較する比較器、(12a)~(12d)は比較器(11a)~(11d)から出る一致出力信号線、(13a)~(13d)はAA(4)から読み出された情報をセットするためのAA

データレジスタ、(14a)~(14d)はAAデータレジスタ(13a)~(13d)からのAAデータレジスタ出力信号線、(15a)~(15d)はAAデータレジスタ(13a)~(13d)内のSフラグから出力される信号のためのSフラグ信号線、(16)はSフラグ信号線(15a)~(15d)上の信号を検知するためのSフラグ検知回路、(17)はSフラグ検知回路(16)からのSフラグ制御線、(18)はSフラグ制御線(17)によりAAデータレジスタ出力信号線(14a)~(14d)上のいずれかの信号を選択するためのセレクト、(19)はセレクト⁽¹⁸⁾からのセレクト出力線、(20)はセレクト出力線(19)上のデータを保持するスタンバイレジスタ(SBR)である。

なお、こゝでは、説明の便宜上、AA(4)は4個の列A、B、C、Dから構成されているとする。

次に、このような構成をもつこの発明の実施例について、その動作を説明する。任意所望のプログラムが実行されていくにつれて、AA(4)がアクセスされ、その各列(A~D)の内容(エントリ)がAA出力信号線(8a)~(8d)に読み出されて、

当該エントリ中のPIDとCPID出力信号線(10)の現在実行中のプログラム識別子(CPID)とが比較器(11a)~(11d)によつて比較される。この比較の結果として、一致出力信号線(12a)~(12d)上に一致信号が出力された場合、対応するAAレジスタ(13a)~(13d)内のSフラグをリセット($S=0$)する。これに対して、一致信号(12a)~(12d)が出力されない場合は、対応するAAデータレジスタ(13a)~(13d)内のSフラグをセット($S=1$)する。この場合、エントリのVフラグがセットされていない($V=0$)ときには、Sフラグのセット、リセット操作は行われない。

このようにしてSフラグの所要のセット、リセット操作が完了すると、この状態を表わす信号はSフラグ信号線(15a)~(15d)経由でSフラグ検知回路(16)に送られる。Sフラグ検知回路(16)においては、この受入れた信号に基づいて、AAデータレジスタ(13a)~(13d)の中でSフラグがリセットされているものをセレクトする為

の信号を作成し、この信号はSフラグ制御線(17)を介してセレクト(18)に送られる。そして、セレクト(18)においては、AAデータレジスタ出力信号線(14a)~(14d)の中の1本がセレクトされ、セレクトされた信号線上の信号はセレクト出力線(19)によりSBR(20)に送られる。このとき、SBR(20)が動作中でなければ、セレクト出力線(19)上のデータはSBR(20)にセットされる。これと同時に、セットしたSBR(20)の内容と同じエントリ情報のVビットはリセットされ、AAデータ出力線(14a)~(14d)経由でAA(4)に返戻される。これによりSBR(20)にセットされたエントリは無効なものとなり、次にブロック交換の必要が生じた場合に利用されることになる。

そして、SBR(20)にセットされたデータに対応するデータがHSB(2)から読み出されて、MM(J)に転送される。この場合、転送動作はBXU(1)からHSB(2)へのアクセス動作とは関係し並行して実行される。また、SBR(20)が動作中の場合には、AAデータレジスタ出力線(14a)~(14d)経由

でそのままAA(4)に書戻される。そのままAA(4)に書き戻されたエントリ情報はSBR(20)の動作が完了し次の追い出し動作が可能になるまでV=1, S=1としてAA(4)に存在することになる。

そして、V=1, S=1の状態にあるエントリはSBR(20)の動作が完了した時点で次に実行されることになる。

なお、SBR(20)の動作とEXU(1)からのアクセス動作とは独立して行われる。また、上記実施例ではAAの構成を4列のものとして説明されたが、これに限られるものではない。

(発明の効果)

以上のように、この発明によればAAのエントリにプログラム識別子、Sフラグを設定し、キャッシュアクセスの空き時間に不要となつたブロックデータを主メモリ装置側に追い出しておくように構成したので、ブロック交換動作時のオーバーヘッドが少なくなり従つてキャッシュまたはバッファ・メモリ装置の使用効率が高まり、更には情報処理システム全体の効率が向上するという効果が

ある。

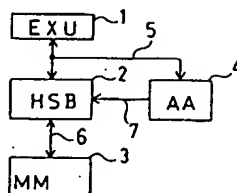
図面の簡単な説明

第1図は一般的なバッファ・メモリ装置を示すブロック図、第2図はこの発明の実施例において用いられるアドレス・アレイのエントリ構成を表わすフォーマット図、第3図はこの発明の実施例の要部構成を示すブロック図である。

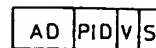
(1)・・・EXU、(2)・・・HSB、(3)・・・MM、(4)・・・AA、(5)・・・HSBバス線、(6)・・・MMバス線、(7)・・・HSBアドレス線、(8a)～(8d)・・・AA出力信号線、(9)・・・CPIDレジスタ、(10)・・・CPID出力信号線、(11a)～(11d)・・・比較器、(12a)～(12d)・・・一致出力信号線、(13a)～(13d)・・・AAデータレジスタ、(14a)～(14d)・・・AAデータレジスタ出力信号線、(15a)～(15d)・・・Sフラグ信号線、(16)・・・Sフラグ検知回路、(17)・・・Sフラグ制御線、(18)・・・セレクト、(19)・・・セレクト出力線、(20)・・・SBR。

なお、各図中、同一符号は同一、又は相当部分を示す。

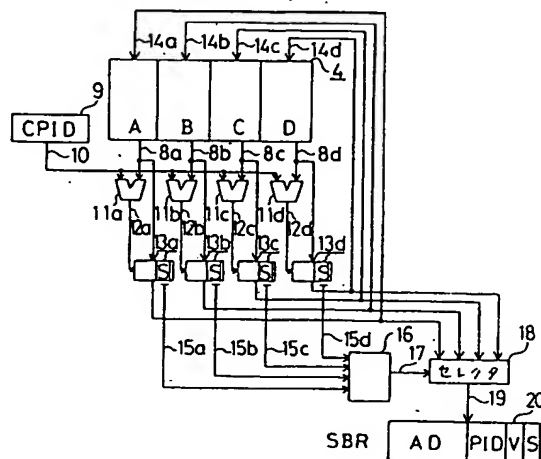
第1図



第2図



第3図



(11) Publication number : Japanese Patent Laid-Open No. 51-19453
(43) Date of publication of application : February 16, 1976
(54) Title of Invention : BUFFER MEMORY CONTROL METHOD

1. Title of the Invention

BUFFER MEMORY CONTROL METHOD

2. Claims

A buffer memory control method

wherein, in a data processing system having a buffer memory and priority managing means which manages priorities of the block units with respect to a predetermined number of block units transferred onto the buffer memory, information for managing, with respect to pieces of block unit information transferred onto the buffer memory, validity of the pieces of block unit information is set, the lowest priority is given to invalid block unit information by the priority managing means on the basis of the information, and information which designates that the block unit information is not invalid but should be replaced with respect to the block unit transferred onto the buffer memory is set, so that control is performed by the priority managing means so that a lower priority is given to a block unit with the information.

3. Detailed Description of the Invention

The present invention relates to a buffer memory control method, and particularly to a buffer memory control method in which, in a replacing process which transfers a block unit transferred onto a buffer memory, information which designates that predetermined block unit information is not valid and information which designates that the block unit information is valid but should be replaced in comparison with other block unit information is held, and control is performed to make it easy to positively replace a block

unit to which the former information or the latter information is given.

In a data processing device having a buffer memory, since the number of block units which can be stored on the buffer memory is limited, control is performed such that a priority order managing means is arranged to determine a next replaceable block unit each time the buffer memory is accessed.

In this case, in a conventional technique, according to the LRU (Least Recently Used) logic, the highest priority is given to a block unit most recently used, and control is performed such that a block unit which is the opposite of the least recently used block unit is set as a block unit to be replaced. Even in a method which does not employ the LRU logic, a conventional logic which determines a block unit to be replaced depends on a passive logic, and a measure which determines a block unit which is more likely to be replaced than the other block units is employed. In contrast to the passive logic, as an active logic, a block unit which is apparently invalid is merely designated.

It is an object of the present invention to designate a block unit which is not invalid as information but should be positively replaced and to positively replace the block unit by a priority order managing means. For this purpose, according to the present invention, there is provided a buffer memory control method wherein, in a data processing system having a buffer memory and a priority managing means which manages priorities of the block units with respect to a predetermined number of block units transferred onto the buffer memory, information for managing, with respect to pieces of block unit information transferred onto the buffer memory, validity of the pieces of block unit information is set, the lowest priority is given to invalid block unit information by the priority managing means on the basis of the information, and information which designates that the block unit information is not invalid but

should be replaced with respect to the block unit transferred onto the buffer memory is set, so that control is performed by the priority managing means so that a lower priority is given to a block unit with the information. Hereinafter, description shall be made with reference to the diagrams.

FIG. 1 shows an embodiment of a data processing system to which the buffer memory control method according to the present invention is applied. FIG. 2 shows a configuration of address information for designating specific stored information in the embodiment. FIG. 3 shows a configuration of replaced block unit determination according to the embodiment of the present invention.

In FIG. 1, reference numeral 1 denotes a main storage device; 2, a central processing device having a buffer memory; 3, a buffer memory; 4, a directory which manages address information of a block unit stored on the buffer memory 3; 5, a memory access device; 6, a command control unit; 7, an arithmetic processing unit; 8-0 and 8-1, channel control devices; 9, a file control device; 10, a large-capacity external storage device; 11, an input/output control device; and 12, an input/output device.

The main storage device 1 is partitioned by block units 13 each constituted by, for example, 8 words, and has a total of m (tags) \times n (columns) block units. On the buffer memory 3, for example, four block units can be transferred and stored every column. When a need arises for accessing information in a block unit which is not present on the buffer memory 3, control is performed to replace a block unit having the lowest priority of the four blocks. More specifically, in response to an access from an access source represented by the central processing device 2 or the channel control device 8-0 or 8-1, the directory 4 is indexed and when necessary information is present on the buffer memory 3, the corresponding information is read from the buffer memory 3. When

necessary information is not present on the buffer memory 3, the main storage device 1 is accessed to transfer the necessary information to the access source, and a block unit including the information is stored on the buffer memory 3. At this time, on the buffer memory 3, a block unit having the lowest priority of the four block units is driven out.

In this case, as the logic for determining a block to be replaced, conventionally, an LRU logic is commonly employed. Furthermore, when a block unit information present on the buffer memory 3 becomes invalid since the block unit information is replaced by another device in the main storage device 1, the invalid block unit is replaced.

According to the present invention, in addition to the conventional replacing process logic, a block unit to be replaced is actively designated. More specifically,

(1) As shown in FIG. 1, when the channel control device 8-0 transfers data of the main storage device to the large-capacity external storage device 10, information of a block unit including the data is stored on the buffer memory 3. However, when considering the case after the last data in the block unit is accessed, the presence of such block unit on the buffer memory 3 is no longer of any value.

(2) When information accessed by the central processing device 2 and information accessed by the channel control device 8-0 or 8-1 are considered in comparison with each other, a probability of sequentially accessing adjacent addresses depending on the progress of the process is high in the former. In the latter, the probability is low. For this reason, the value of the presence of the block unit on the buffer memory 3 after the latter access is low.

(3) When access by an access source is considered, the value of the presence on the buffer memory 3, for an access corresponding to a user area, for example, is low in comparison to others.

(4) With respect to a program in the data processing device, in consideration of a monitor mode and a user mode, the value of the presence on the buffer memory 3 for the latter is low.

(5) Depending on a type of a process of the data processing device, there are cases where it may be predicted that for a certain period of time after information is accessed, the probability of accessing the block unit including the information again is nil or very low. In such a case, once a specific block unit is accessed, the presence of the block unit on the buffer memory 3 is no longer of any value.

In these various cases, a block unit whose presence on the buffer memory 3 has no value or less value is designated as a block unit which should be actively replaced automatically or by a command.

The address information according to the present invention, as expressed in FIG. 2, is designated by (tag · address information), (column · address information), and (in-block address information).

In FIG. 3, reference numeral 4 denotes a directory; 13, an address register; 14-0, 14-1, 14-2, and 14-3 denote comparing circuits; 15, a priority determining circuit; and 16-0, 16-1, 16-2, and 16-3, lowest order set flip-flop.

As is explained in relation to FIG. 1, when the buffer memory 3 is accessed, the access source sets address information in the block units 13 and performs a reading process on the directory 4 based on column address information. When the column address information corresponds to, for example, #0 column, pieces of tag address information are simultaneously read from four set positions, i.e., a #0 set position to a #3 set position, of the #0 column in the directory 4. The pieces of read tag address information are pieces of tag address information in block units stored at four set positions of the #0 column of the buffer memory 3 (FIG. 1), and are compared with the tag address information on the register 13 by the comparing circuits 14-0 to 14-3. When non-match signals are

generated from all the comparing circuits, a flip-flop, for example, the flip-flop 16-1 in a set state drives out a block unit stored at a corresponding set position on the buffer memory 3, i.e., the #1 set position. Then, a block unit is transferred from the main storage device 1 to such position. The highest priority is given to the #1 set position at which the new block unit is stored, the lowest order is newly determined, and the corresponding flip-flop, for example, the flip-flop 16-2 is set.

In this case, pieces of information V which designate whether a block unit stored on the buffer memory 3 is valid information are stored in the directory 4 at the same time. In a determination of the lowest priority, when a block unit having the information V being logical "0" (i.e., invalid) is present, a flip-flop at a set position where the block unit is present is set.

In the present invention, in addition to the priority determining process using the information V and, for example, the LRU logic, information, shown as information R (not limited to 1 bit), which designates that "information is not valid but should be replaced" is stored on the directory 4, and a block unit in which the information R is logical "1" (i.e., should be replaced) is actively replaced. More specifically, first, the lowest priority is given to a block in which the information V is logical "0". Secondly, the lowest priority is given to a block in which the information R is logical "1" when all the pieces of information V are logical "1". Thirdly, when the pieces of information V are logical "1" and the pieces of information R are logical "1" in all the blocks, the lowest-priority block unit is determined according to, for example, the LRU logic.

Whichever block unit has the information R being logical "1" is determined according to the aforementioned item (1) or (5) or the like. A process of changing the information R into information being logical "1" is performed automatically or by a command.

As described above, according to the present invention, in a

process for determining the block to be replaced performed by the priority determining circuit 15, a block to be actively replaced is instead processed, using the information R, so as to have a form which can be more easily replaced, and buffer memory control can be performed more efficiently than in a method in which, as in the conventional process of this type, a logic such as the LRU logic is applied such that all the block units which are not invalid information are equally regarded.